SRM VALLIAMMAI ENGINEERING COLLEGE
(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF INFORMATION TECHNOLOGY

M. Tech- Data Science

QUESTION BANK

II SEMESTER

DS3261 DATA PREPROCESSING AND OPTIMIZATION TECHNIQUES

Regulation – 2023

Academic Year 2024 – 2025(Even Semester)

Prepared by

Ms Abinaya J Assistant.Professor (o.G) /IT

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF INFORAMTION TECHNOLOGY

QUESTION BANK

SUBJECT : DS3261– DATA PREPROCESSING AND OPTIMIZATION TECHNIQUES
SEM/YEAR : II / I

UNIT – I INTRODUCTION TO DATA PREPARATION			
Data Sets and Proper Statistical Analysis of Data Mining Techniques-Data Preparation Basic Models-Data Integration-Data Cleaning-Data Normalization			
PART – A			
Q. No	**Question**	**BTL**	**Competence**
1	Define Data mining.	BTL 1	Remembering
2	Analyze the conditions are needed in order to safely carry out parametric.	BTL 2	Understanding
3	What is a data set?	BTL 5	Remembering
4	What is the significance of the validation set during model training?	BTL 3	Applying
5	Identify the purpose Iman–Davenport test.	BTL 1	Remembering
6	Why is data cleaning needed?	BTL 4	Analyzing
7	Examine data integration.	BTL 4	Analyzing
8	Mention the purpose of KEEL DM tool	BTL 4	Analyzing
9	Define data cleaning.	BTL 1	Remembering
10	Mention the formula for z-score normalization.	BTL 4	Analyzing
11	Write the formula for Min-max normalization.	BTL 3	Applying
12	Illustrate the various process in Data Integration.	BTL 5	Evaluating
13	What is Decimal Scaling Normalization.	BTL 4	Analyzing
14	Illustrate the Underfitting problem	BTL 5	Evaluating
15	Define Overfitting data problem.	BTL 6	Creating
16	Summaries the uses of . post-hoc tests.	BTL 2	Understanding
17	Analyse the Data Transformation Strategies	BTL 2	Understanding
18	Analyze the Friedman Test and Post-hoc Tests.	BTL 5	Evaluating
19	What is the Chi-square ($\chi 2$) correlation test.	BTL 6	Creating
20	Why is data pre-processing important?	BTL 1	Remembering
21	Write the steps for decimal scaling normalization.	BTL 5	Evaluating
22.	Differentiate between parametric and non-parametric tests	BTL 3	Applying
23.	Measure the major tasks in data preprocessing.	BTL 1	Remembering
24	What is Z-score Normalization.	BTL 5	Evaluating

PART - B				
Q. No	**Question**	**Marks**	**BTL**	**Competence**
1	Discuss in detail about (i) Friedman Test and Post-hoc Tests.	07	BTL 4	Analyzing

	(ii) Iman–Davenport test	08		
2	What is tuple duplication in the context of a dataset, and why is it important to detect and handle it?	15	BTL 6	Creating
3	Discuss issues to consider during data integration.	15	BTL 2	Understanding
4	Criticize the value ranges of the following normalization methods. i. Z-Score normalization ii. Normalization by decimal scaling	07 08	BTL 5	Evaluating
5	Describe the K-Fold Cross validation process.	15	BTL 4	Analyzing
6	Explain in detail about Min-Max Normalization with an example	15	BTL 4	Analyzing
7	Discuss in detail about Data Set Partitioning process.	15	BTL 2	Understanding
8	What is data normalization, and why is it important in data analysis and machine learning?	15	BTL 1	Remembering
9	What is data cleaning, and why is it important in the context of data analysis process.	15	BTL 3	Applying
10	Describe the process of conducting a Chi-square ($\chi 2$) correlation test. What are the key steps involved?	15	BTL 3	Applying
11	Discuss in detail about Data Normalization process.	15	BTL 2	Understanding
12	Explain in detail about Decimal scaling normalization and calculate decimal scaling normalization for the given dataset $X=\{-5,0,23.0,17.6,9.2,3.1,11\}$:	15	BTL 4	Analyzing
13	Describe two common methods for detecting redundant attributes in a dataset. How do these methods work?	15	BTL 4	Analyzing
14	Discuss in detail about Data Normalization process.	15	BTL 1	Remembering
15	Calculate Z-score normalization for the dataset $X=\{-5,0,23.0,17.6,9.2,3.1,11\}$ scales the data so that it has a mean of 0 and a standard deviation of 1.	15	BTL 3	Applying
16	Explain the concept of partitioning a data set into training, validation, and testing sets. Why is this partitioning necessary?	15	BTL 3	Applying
17.	Perform min-max normalization on the dataset $X=\{-5,0,23.0,17.6,9.2,3.1,11\}$ to scale the values within the interval [0,1]	15	BTL 4	Analyzing

UNIT – II DATA PRE-PROCESSING & TRANSFORMATION

Dealing with Missing Values-Dealing with Noisy Data-Linear Transformations-Quadratic Transformations-Non-polynomi Approximations of Transformations-Polynomial Approximations of Transformations-Nominal to Binary Transformation.

PART – A

Q. No	Question	BTL	Competence
1	How nominal to binary transformation works?	BTL 2	Understanding
2	Evaluate *Iterative-Partitioning Filter*	BTL 4	Analyzing
3	Mention some common non-polynomial approximations of transformations	BTL 1	Remembering
4	Define Attribute noise.	BTL 5	Evaluating
5	Mention the uses of Quadratic transformations	BTL 4	Analyzing
6	Describe Linear transformation.	BTL 1	Remembering
7	Define Class noise	BTL 1	Remembering
8	List out the Types of Noise Data	BTL 2	Understanding
9	Identify various types of noise data.	BTL 3	Applying
10	List the Characteristics of Noise Generation.	BTL 4	Analyzing
11	List out the some common non-polynomial approximations of transformations.	BTL 5	Evaluating
12	Summarize the types of data Transformation.	BTL 6	Creating
13	What is Expectation-Maximization (EM)	BTL 2	Understanding
14	How to apply Data polishing methods?	BTL 3	Applying
15	What is Noise Filter?	BTL 1	Remembering
16	Illustrate Ensemble Filter	BTL 2	Understanding

17	How to filtering the noise by using Cross-Validated Committees Filter?	BTL 3	Applying
18	How Iterative-Partitioning Filter differ from cross validated committees filter?	BTL 3	Applying
19	Recall Robust learners.	BTL 1	Remembering
20	Mention Three types of problems are usually associated with Missing values	BTL 4	Analyzing
21	Analyze benefits of Data Transformation	BTL 4	Analyzing
22	Evalute Data polishing methods.	BTL 5	Evaluating
23	Illustrate various methods of data Transformation.	BTL 6	Creating
24	Define Attribute noise.	BTL 2	Understanding

PART – B				
Q. No	**Question**	**Marks**	**BTL**	**Competence**
1	Write a short notes on : i) Linear transformation. ii) Quadratic transformation.	 07 08	BTL 1	Remembering
2	Explain non polynomial approximations of transformation .	15	BTL 5	Evaluating
3	Explain polynomial approximations of transformation .	15	BTL 5	Evaluating
4	Describe the features of (i) Rank transformation. **(ii)** Box-Cox Transformations.	 07 08	BTL 6	Creating
5	How can transformation done from nominal to binary?	15	BTL 3	Applying
6	Describe the treatment of Missing Values in Data Mining.	15	BTL 3	Applying
7	Discuss the effect of the Imputation Methods in the Attributes' Relationships	15	BTL 3	Applying
8	Analyze Polynomial Approximations of data Transformations.	15	BTL 4	Analyzing
9	How to deal with noisy data in accordance with obtaining higher classification accuracies on test data. Justify.	15	BTL 3	Applying
10	Mention various Types of Noise Data and describe each of them.	15	BTL 2	Understanding
11	Explain in detail about Cross-Validated Committees Filter	15	BTL 4	Analyzing
12	Discuss about: (i) Iterative-Partitioning Filter. (ii) Ensemble Filter .	 07 08	BTL 1	Remembering
13	Illustrate the various techniques for handling noisy data.	15	BTL 1	Remembering
14	Discuss in detail about Bayesian Principal Component Analysis (BPCA).	15	BTL 2	Understanding
15	(i) Explain the various types of Normalization. (ii) Discuss the various benefits and challenges of Data Transformation	07 08	BTL 4	Analyzing
16	Analyze the common techniques for handling missing values.	15	BTL 4	Analyzing
17	Analyze the Non-polynomial Approximations of Transformations with an own example.	15	BTL 4	Analyzing

UNIT - III: DATA REDUCTION &FEATURE SELECTION			
Principal Components Analysis-Data Sampling-Data condensation, data squashing, data clustering, Feature Selection-Exhaustive Methods-Heuristic Methods-Nondeterministic Methods-Instance Selection: Prototype Selection Taxonomy-Edition Algorithms -Hybrid Algorithms.			
PART – A			
Q. No	**Question**	**BTL**	**Competence**
1	Illustrate the concept of data reduction.	BTL 2	Understanding
2	Define Data Sampling.	BTL 1	Remembering
3	What you meant by data condensation?	BTL 3	Applying
4	Define Principal Component Analysis (PCA).	BTL 4	Analyzing
5	List out the benefits of applying data squashing to a dataset	BTL 5	Evaluating
6	What is likelihood-based data squashing?	BTL 6	Creating

7	How does data squashing contribute to data analysis?	BTL 3	Applying
8	How does data condensation differ from data compression?	BTL 2	Understanding
9	What is data sampling, and why is it important in data analysis?	BTL 4	Analyzing
10	How would you select a representative sample from a large dataset.	BTL 3	Applying
11	What is data clustering, and what are its primary objectives in machine learning and data analysis?	BTL 2	Understanding
12.	Analyze the few possible interpretations where "data squashing" might be used.	BTL 5	Evaluating
13.	What are the common techniques used for feature selection?	BTL 4	Analyzing
14.	Analyze the filter model properties.	BTL 1	Remembering
15.	What are the Various limitations of wrapper feature selection techniques?	BTL 2	Understanding
16.	What is embedded feature selection, and how does it differ from filter and wrapper approaches?	BTL 1	Remembering
17	What is the objective of instance selection, and how does it differ from feature selection	BTL 3	Applying
18.	What is Repeated-ENN (RENN).	BTL 1	Remembering
19.	Mention the Categories the algorithms used for Instance selection.	BTL 1	Remembering
20.	Write the purpose of Condensed Nearest Neighbor (CNN) Algorithm.	BTL 2	Understanding
21	What are heuristic methods in feature selection, and how do they differ from exhaustive approaches.	BTL 6	Creating
22	What is Condensed Nearest Neighbor (CNN), and what role does it play in TCNN?	BTL 3	Applying
23	How do you evaluate the quality of prototypes selected through clustering	BTL 2	Understanding
24	Mention the purpose of hybrid algorithm in instance selection.	BTL 4	Analyzing

PART – B

Q. No	Question	Marks	BTL	Competence
1	What are the different methods of data sampling, and when would you use each method?	15	BTL 2	Understanding
2	(i).What is Principal Component Analysis (PCA), and what is its primary objective in data analysis? (ii) Describe basic procedure to find correlation in Principal Component Analysis.	8 7	BTL 6	Creating
3	How does PCA help in dimensionality reduction, and why is dimensionality reduction important in data analysis?	15	BTL 3	Applying
4	What are the steps involved in performing PCA on a dataset	15	BTL 6	Creating
5	Discuss: (i) Edition Algorithms. (ii) Hybrid Algorithms.	8 7	BTL 1	Remembering
6	(i).What are the advantages and limitations of using sampling for data condensation? (ii) Difference between supervised and unsupervised learning, and where does clustering fit in.	8 7	BTL 2	Understanding
7	What are some examples of common algorithms or techniques used in nondeterministic feature selection.	15	BTL 5	Evaluating
8	(i).How do you determine the optimal subset of features in Exhaustive Methods. (ii).What are some common algorithms or techniques used in Exhaustive Methods?	8 7	BTL 4	Analyzing

9	Explain the process of Exhaustive Methods in detail	15	BTL 3	Applying
10	How do you evaluate the performance of nondeterministic methods in feature selection	15	BTL 2	Understanding
11	Analyze the various algorithms used for instance selection.	15	BTL 1	Remembering
12	(i).Write the main reasons for performing Feature Selection. (ii) Explain the concept of filter methods and their role in feature selection	8 7	BTL 1	Remembering
13	Illustrate the Various approaches used to select the most relevant features from a dataset.	15	BTL 1	Remembering
14	What are some common criteria used in edition algorithms for selecting or discarding instances?	15	BTL 4	Analyzing
15	Discuss: (i) Embedded Feature Selection (ii) Wrappers	8 7	BTL 1	Remembering
16	Analyze the various Criteria to Compare Prototype Selection Methods.	15	BTL 4	Analyzing
17	Discuss the Common Properties in Prototype Selection Methods	15	BTL 1	Remembering

UNIT - IV: UNCONSTRAINED OPTIMIZATION

Introduction-Conditions for Local Minimizes-Golden Section Search-Newton's Method-Secant Method- Gradient Methods-Steepest Descent-Conjugate Direction Methods-Unconstrained Optimization and Neural Networks.

PART – A

Q. No	Question	BTL	Competence
1	Define Optimization.	BTL 1	Remembering
2	Differentiate constrained and unconstrained optimization	BTL 4	Analyzing
3	What is the main principle behind the Golden Section Search algorithm?	BTL 1	Remembering
4	Explain unconstrained optimization.	BTL 1	Remembering
5	When the Gradient method is useful?	BTL 3	Applying
6	Define Newton method.	BTL 2	Understanding
7	Drive hessian Matrix	BTL 5	Evaluating
8	Derive the expression for activation function.	BTL 5	Evaluating
9	Define artificial neural network	BTL 2	Understanding
10	How constraints is important for optimization?	BTL 4	Analyzing
11	Write down the condition for local minimizers	BTL 3	Applying
12	Summarize objective of Golden Search Method.	BTL 2	Understanding
13	List out the one-dimensional search methods	BTL 1	Remembering
14	List the properties of conjugate direction.	BTL 1	Remembering
15	How gradient involved in conjugate direction method?	BTL 4	Analyzing
16	Draw the structure of feedforward neural network	BTL 6	Creating
17	Define Back propagation algorithm	BTL 2	Understanding
18	Write the properties of conjugate direction method.	BTL6	Creating
19	Construct the artificial neural network for single neuron	BTL 6	Creating
20	Define Saddle point	BTL 2	Understanding
21	Why conjugate gradient is better than steepest descent?	BTL 2	Understanding
22	Define: level set.	BTL 2	Understanding

23	Compare First and Second order condition in optimization?	BTL 6	Creating
24	Illustrate of conjugate direction.	BTL6	Creating

PART - B				
Q. No	**Question**	**Marks**	**BTL**	**Competence**
1	Describe Optimization technique and Explain constrained and unconstrained optimizations	15	BTL 1	Remembering
2	Discuss the Conjugate Direction Algorithm with theorem.	15	BTL 4	Analyzing
3	Explain in detail about conjugate direction algorithm.	15	BTL 1	Remembering
4	Use the golden section search method to find the value of x, which minimize $f(x)=x^2-6x+20$. The interval to be considered is x=0 and x=10.	15	BTL 5	Evaluating
5	Identify the Conditions for Local Minimizers and Explain with example.	15	BTL 1	Remembering
6	How Golden Section Search suitable for minimizer and explain with suitable example	15	BTL 3	Applying
7	(i)Write the relationship between gradient and level curves? (ii) How is Taylor series used in deep learning?	08 07	BTL 5	Evaluating
8	Describe Newton method and how it can be applied in Optimization.	15	BTL 3	Applying
9	Write about Conjugate direction method	15	BTL 5	Evaluating
10	Define one dimensional search method. Explain any one of the one-dimensional search methods.	15	BTL 1	Remembering
11	Illustrate the process of back propagation algorithm in three layered neural network	15	BTL 2	Understanding
12	Illustrate Golden Section Search with example	15	BTL 2	Understanding
13	(i).Define Optimization. Why we need Optimization for Machine learning? (ii).Necessary and Sufficient optimality conditions for optimization problems.	08 07	BTL 4	Analyzing
14	Explain how to minimize the function f of a single real variable x using newton method?	15	BTL 4	Analyzing
15	Discuss the Method of Steepest Descent with example.	15	BTL 5	Evaluating
16	Explain Golden-section search for optimization in 1-D.	15	BTL 3	Applying
17	Mention procedure to update the weights in back propagation network and how it's optimized.	15	BTL 5	Evaluating

UNIT - V: CONSTRAINED OPTIMIZATION			
Simplex Method - Nonsimplex Methods - Problems with Equality Constraints : Tangent and Normal Spaces-Lagrange Condition-Convex Optimization Problems Projections-Projected Gradient Methods with Linear Constraints.			
PART – A			
Q. No	**Question**	**BTL**	**Competence**
1	Mention the importance of constrained optimization	BTL 2	Understanding
2	Define pivot element.	BTL 1	Remembering
3	Describe the Objective of optimization.	BTL 2	Understanding
4	Enumerate the difference between Constrained and Unconstrained Optimization	BTL 5	Evaluating
5	Infer difference between local and global minimizer	BTL 5	Evaluating
6	What is a projected gradient method, and how does it work?	BTL 1	Remembering
7	Mention the two methods of solving Linear programming problem	BTL 2	Understanding

8	Write an example for constrained optimization.	BTL 6	Creating
9	What is the Lagrangian function, and how is it defined?	BTL 1	Remembering
10	Compare Graph and epigraph	BTL 5	Evaluating
11	How do you choose the step size (or learning rate) in a projected gradient method?	BTL 2	Understanding
12	What are some convergence properties of projected gradient methods?	BTL 4	Analyzing
13	Write Karush-Kuhn-Tucker condition	BTL 3	Applying
14	How penalty method used for optimization?	BTL 4	Analyzing
15	State the significance of Lagrange's theorem in optimization	BTL 2	Understanding
16	Give some examples of application of optimization concepts.	BTL 4	Analyzing
17	Drive tangent plane and space	BTL 3	Applying
18	State Lagrange's Theorem	BTL 1	Remembering
19	What are the Lagrange conditions in optimization?	BTL 3	Applying
20	State nonlinear programming problems.	BTL 1	Remembering
21	Compare hard and soft constraints.	BTL 3	Applying
22	What are the different types of penalty functions?	BTL 3	Applying
23	What are the goals and constraints of linear optimization?	BTL 3	Applying
24	When constraints hold with equality?	BTL 3	Applying

PART - B

Q. No	Question	Marks	BTL	Competence
1	What are nonsimplex optimization methods, and how do they differ from simplex methods?	15	BTL 6	Creating
2	Derive matrix form of the simplex method.	15	BTL 3	Applying
3	Explain procedure for two phase simplex method	15	BTL 2	Understanding
4	How do you Solve the constrained problem by projected gradient descent?	15	BTL 5	Evaluating
5	Explain simplex method of solving linear equations with an example	15	BTL 2	Understanding
6	Write the basic approach of the Penalty function method.	15	BTL 5	Evaluating
7	Consider the following linear programming problem and maximize $7x1+ 6x2$ with subject to $2x1 + X2 \leq 3$ $x1 + 4x2 \leq 4$, $x1,x2>=0$	15	BTL 6	Creating
8	Explain Khachiyan's Method in detail.	15	BTL 1	Remembering
9	Illustrate Lagrange's Algorithm for Equality and Inequality constraints.	15	BTL 4	Analyzing
10	How can address optimization problems with equality constraints.	15	BTL 1	Remembering
11	Explain Convex Optimization Problems or programming	15	BTL 2	Understanding
12	Explain the concept of Lagrange's Algorithm for Inequality constraints.	15	BTL 1	Remembering
13	How Projected Gradient Methods carried out with Linear Constraints	15	BTL 3	Applying
14	Explain Lagrangian Algorithm for Equality Constraints	15	BTL 2	Understanding
15	Describe the detail procedure for simplex method with example.	15	BTL 3	Applying
16	Explain in detail the Sequential linear programming method.	15	BTL 3	Applying
17	Discuss the Equality and Inequality Constraints in Optimum Design Problem Formulation.	15	BTL 3	Applying

VALLIAMMAI ENGINEERING COLLEGE
SRM